

Summary of XMP explorations with external contacts

DRAFT – Caroline Arms. 2011-12-09

Digital Library Community

Contacted: Harvard, MIT, Stanford, Virginia Tech, Michigan, California Digital Libraries, DuraSpace,

Essentially no use of XMP was found. Embedded metadata doesn't solve any of the problems associated with digital repositories. In several instances some consideration had been given to using XMP, with explicit decisions made against pursuing it. In this context metadata for discovery is usually managed outside the content objects, where it can be modified and improved without affecting the underlying object. Embedded metadata is hard to manage. Some institutions had considered embedding XMP in dynamically created derivatives, e.g., for pictorial images, but had not pursued the idea. The only use of XMP identified was by Stanford, where they embed rights information (using a Creative Commons license) into theses.

Some institutions reported that they knew a few PDFs ingested into an institutional repository contained XMP, but they had made no use of it, because it was so seldom present. If present, it probably only replicated information in the Basic PDF metadata fields.

Scholarly Publishing Community

Contacted: Nature, CrossRef, Portico,

Not much use of XMP in published PDFs. No shared tools or guidelines. A new service from CrossRef, CrossMark, may stimulate more XMP adoption in a "standard" approach, but it remains to be seen if CrossMark is picked up by publishers. See <http://www.crossref.org/crossmark/index.html>

Nature has been embedding XMP with PDFs of articles for 2 or 3 years, and reported that Elsevier had also been using XMP for some publications. The two publishers each chose their own set of elements and content practices for those elements. The element sets have some overlap. Portico reports that they have not been able to make any effective use of embedded XMP. It is seldom present and if present, it reflects different practices by publisher and lack of consistency within a publisher's titles and over time. For ingest of PDFs into Portico, they ask for (and have no problem in getting) separate article metadata in XML. This is typically exported dynamically from an asset management system and is the metadata that is actively used in the publishing process, and hence likely to get corrected if in error.

Digital Photography and Still Image Community

Contacted: Greg Reser (UCSD, VRA)

The digital photography community has adopted XMP to a great degree, to hold IPTC and PLUS metadata, and combat the "orphan works" problem. However, tools associated with image manipulation and management software are not of use for textual PDFs. The metadata namespaces used by this community (IPTC, PLUS) are unlikely to be convenient to use for LC purposes without development work to collect the relevant elements into a single view for data entry.

Greg Reser of UCSD has worked under the auspices of VRA to develop data entry capabilities for use with Adobe Photoshop to embed elements from the VRA Core metadata set. His work and experience is the most relevant I have found to LC, since it is possible to use the same mechanisms in other Adobe tools. Considerable development and testing would be needed to develop something similar for LC use with PDFs. [Aside: Reser would be very enthusiastic if LC chose to be active in this area on behalf of the broader community, but he is only interested in images. P&P has not endorsed or used VRA so this does not necessarily seem a useful direction to encourage at this point. However, he might be able to give an interesting talk at LC about the work he has done.]

Niche use of XMP for Creative Commons

The Creative Commons community has developed some tools to make it straightforward to embed rights information into PDFs and image formats. See, for example, <http://www.adobetutorialz.com/articles/159/1/Creative-Commons-Metadata-and-XMP> and <http://wiki.creativecommons.org/XMP>